

REMARKS

Claims 1-9 are all the claims pending in the application; each of the claims has been rejected.

Claims 1-7 are being canceled and replaced by claims 10-16, respectively. Claims 8 and 9 are also being cancelled, but no replacement claims are being submitted in their place.

New claims 10-16 recite the same subject matter as canceled claims 1-7. The new claims more fully use U.S. format and grammatical errors in the claims have been corrected. As explained below, additional amendments to the wording has been made to more clearly recite that which Applicants regard as their invention.

No new matter has been added. Entry of the amendment is respectfully requested.

I. Rejection of Claims Under 35 U.S.C. §112

At paragraph 4 of the Office Action, claims 1-9 are rejected under 35 U.S.C. §112, second paragraph, as being indefinite.

The Examiner states that claim 1 is indefinite because it is unclear what is intended by the phrase “specific region” in line 6. The Examiner requests clarification.

In response, Applicants note that the present invention is a method for determining whether a “specific region” of DNA from the genome of an organism (i.e., a selected stretch of polynucleotides) encodes a polypeptide or a peptide fragment, based on the existence of corresponding mRNA. As not all of the DNA in an organism’s genome encodes proteins, some of the DNA in an organism’s genome is not expressed (e.g., introns and so-called junk DNA). It is the expressed DNA or the exons (or portions thereof) that are the “specific regions” referred to in Applicants’ application.

The invention does not determine whether a particular piece of DNA exists, but rather the invention determines whether a particular piece of DNA encodes a polypeptide or a peptide fragment that is expressed in a cell, based on the presence of corresponding mRNA. Indeed, the sequence of the particular piece of DNA is known prior to conducting the method recited in the claims.

In order to more clearly recite that which Applicants regard as their invention, claims 1-9 have been canceled and new claims 10-16 are presented in their place. The subject matter recited in the new claims corresponds in numerical order to the subject matter recited in the canceled claims (as claims 8 and 9 are being canceled in view of the §§102 and 103 rejections, new claims corresponding to these two claims are not being presented).

For the sake of clarity, the term “selected DNA molecule” is used in the new claims to substitute for the term “specific region” in the canceled claims. Applicants also note that the term “gene expression region” is used in the claims. The specification makes it clear that a “gene expression region” is a region of DNA that corresponds to a gene that is expressed (transcribed and translated) as a protein or fragment thereof. It is any portion of the gene, and is not limited to a specific length. As explained at page 8 of the specification, the length of the DNA molecule (“specific region”) is not particularly limited.

In view of the amendments to the claims (i.e., presentation of new claims), Applicants assert that each of the claims is definite as written. Therefore, Applicants respectfully request reconsideration and withdrawal of this rejection.

II. Rejection of Claims Under 35 U.S.C. §102

At paragraph 7 of the Office Action, claims 1-2 and 7-9 are rejected under 35 U.S.C. §102(b) as being anticipated by Nakamura et al. (USP 5,556,945 (1996)).

The Examiner states that Nakamura et al. teaches a method of determining whether an arbitrary DNA sequence is part of a gene. The Examiner further states that this patent teaches the isolation and structure of a genomic gene which was determined to be a gene expression region by the method of claim 1. The Examiner also states that the patent teaches a protein encoded by the gene, such as that recited in claim 8 of the application.

In response, Applicants note that Nakamura et al. discloses the isolation and cloning of a specific gene using the exon trapping method of Buckler et al. (*PNAS* 88:4005-4009 (1991)). As the patent does not provide sufficient information concerning the method, Applicants also reviewed Buckler et al. which describes the method in detail. For the Examiner's convenience, Applicants enclose herewith a copy of Buckler et al.

A review of the exon trapping method published by Buckler et al. makes it clear that the method is quite different from that of the present application and does not anticipate the claims.

The exon trapping method is summarized in Buckler et al. (page 4006, column 1, sixth full paragraph) as "when a DNA fragment containing an entire exon with flanking intron sequence in the sense orientation is inserted into the *Bam*HI site of the vector [produced for use in the method], the exon should be retained in the mature poly(A)⁺ cytoplasmic RNA."

More precisely, the method utilizes an expression vector engineered to comprise an intron from the HIV-1 *tat* gene, into which was inserted a *Bam*HI restriction site. A segment of DNA which is expected to contain an exon is cloned into the vector at the *Bam*HI site. The

vector is then transfected into COS-7 cells, upon which RNA transcripts are produced. Splice sites of exons contained within the inserted genomic fragment are paired with splice sites of the flanking *tat* intron, and the *tat* intron sequences are spliced out to produce polyadenylated cytoplasmic mRNA molecules containing the coding sequence of the inserted exon. The resulting RNA may then be subjected to reverse transcription to produce cDNA, followed by PCR amplification using primers that correspond to sequences of the vector.

When compared to the method of the instant application, it is clear that the method of Buckler et al. is quite different. First, the method of Buckler et al. physically uses the full length DNA exon undergoing analysis, flanked by 5' and 3' splice sites, in order to confirm that an exon has been isolated from genomic DNA. In contrast, the method of the instant application does not even employ the analysis of a DNA molecule, instead in the present invention it is the RNA transcript that is being analyzed. Only the sequence of the DNA fragment must be known.

Second, the genomic DNA fragment undergoing analysis is cloned into a specialized expression vector in the method of Buckler et al. In contrast, the method of the instant application does not employ cloning of a DNA fragment, nor the expression of a cloned DNA fragment, and it does not use a specialized vector, or any vector for that matter. Indeed, as mentioned above, in the method of the instant application the DNA fragment is not physically isolated, let alone subcloned. Instead, the method only uses RNA transcripts from an organism and the known sequence of the selected DNA molecule, so that appropriate primers can be prepared. Again, the RNA transcript is being analyzed, not the selected DNA molecule.

Third, transfection of the specialized vector containing the DNA fragment to be analyzed into an isolated population of cells, followed by isolation of RNA from these cells, is employed

in the method of Buckler et al. Thus, it is impossible to derive any conclusions concerning which cells naturally express the exon, or the levels of mRNA expression in those cells. In contrast, in the method of the instant application, the mRNA transcripts may be collected from any population of cells, *in vitro* or *in vivo*, for analysis. And again, as explained above, the selected DNA molecule of the present invention is not being analyzed and therefore it is not transfected into a population of cells as in Buckler et al.

For at least these reasons, the exon trapping method of Buckler et al., referenced in the disclosure of Nakamura et al., does not anticipate the instant invention recited in rejected claims 1-2 and 7 (corresponding to claims 10-11 and 16).

Applicants note that claims 8 and 9 have been canceled (no new corresponding claims are being presented). Thus, the rejection with respect to these two claims is moot.

In view of the points discussed above, Applicants respectfully assert that the cited claims are not anticipated by Nakamura et al., and therefore respectfully request reconsideration and withdrawal of this rejection.

III. Rejection of Claims Under 35 U.S.C. §103

At paragraph 10 of the Office Action, claims 8 and 9 are rejected under 35 U.S.C. §102(e) as being anticipated by or, in the alternative, under 35 U.S.C. §103(a) as obvious over Blumenfeld et al. (USP 6,551,792 (2003)).

In response, Applicants note that claims 8 and 9 have been canceled (no new corresponding claims are being presented).

In view of the cancellation of claims 8 and 9, Applicants respectfully assert that the rejection is moot, and therefore respectfully request reconsideration and withdrawal of this rejection.

IV. Conclusion

In view of the above, reconsideration and allowance of this application are now believed to be in order, and such actions are hereby solicited. If any points remain in issue which the Examiner feels may be best resolved through a personal or telephone interview, the Examiner is kindly requested to contact the undersigned at the telephone number listed below.

The USPTO is directed and authorized to charge all required fees, except for the Issue Fee and the Publication Fee, to Deposit Account No. 19-4880. Please also credit any overpayments to said Deposit Account.

Respectfully submitted,



Drew Hissong
Registration No. 44,765

SUGHRUE MION, PLLC
Telephone: (202) 293-7060
Facsimile: (202) 293-7860

WASHINGTON OFFICE

23373

CUSTOMER NUMBER

Date: September 2, 2003

Exon amplification: A strategy to isolate mammalian genes based on RNA splicing

(gene cloning/polymerase chain reaction)

ALAN J. BUCKLER*, DAVID D. CHANG, SHARON L. GRAW, J. DAVID BROOK, DANIEL A. HABER,
PHILLIP A. SHARP, AND DAVID E. HOUSMAN

Center for Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Phillip A. Sharp, January 25, 1991

ABSTRACT We have developed a method, exon amplification, for fast and efficient isolation of coding sequences from complex mammalian genomic DNA. This method is based on the selection of RNA sequences, exons, which are flanked by functional 5' and 3' splice sites. Fragments of cloned genomic DNA are inserted into an intron, which is flanked by 5' and 3' splice sites of the human immunodeficiency virus 1 *tat* gene contained within the plasmid pSPL1. COS-7 cells are transfected with these constructs, and the resulting RNA transcripts are processed *in vivo*. Splice sites of exons contained within the inserted genomic fragment are paired with splice sites of the flanking *tat* intron. The resulting mature RNA contains the previously unidentified exons, which can then be amplified via RNA-based PCR and cloned. Using this method, we have isolated exon sequences from cloned genomic fragments of the murine Na,K-ATPase α_1 -subunit gene. We have also screened randomly selected genomic clones known to be derived from a segment of human chromosome 19 and have isolated exon sequences of the DNA repair gene *ERCC1*. The sensitivity and ease of the exon amplification method permit screening of 20–40 kilobase pairs of genomic DNA in a single transfection. This approach will be extremely useful for rapid identification of mammalian exons and the genes from which they are derived as well as for the generation of chromosomal transcription maps.

Understanding the molecular basis of human genetic disorders and corresponding genotypes in other mammalian model systems requires methods for the identification of coding sequences in target chromosomal regions. Current methods that are used for this purpose are both inefficient and tedious. The strategy used most frequently involves the screening of short genomic DNA segments for sequences that are evolutionarily conserved (1–4). Alternative strategies involve sequencing and analyzing large segments of genomic DNA for the presence of open reading frames (5) and cloning hypomethylated CpG islands, signposts of 5' ends of transcription units (6). However, none of these methods provides a direct means of purifying coding sequences from genomic DNA.

We have developed a method to rapidly and efficiently isolate exon sequences from cloned genomic DNA by virtue of selection for functional 5' and 3' splice sites. Random segments of chromosomal DNA are inserted into an intron present within a mammalian expression vector and, after transfection, cytoplasmic mRNA is screened by PCR amplification for the acquisition of an exon from the genomic fragment. The amplified exon is derived from the pairing of unrelated vector and genomic splice signals. Previous studies have shown that introns constructed with novel combinations of 5' and 3' splice sites from diverse genes are actively spliced

(7, 8). Thus, this method may be generally applicable for the selection of exon sequences from any gene. The method is also both rapid and easily adapted to large scale experiments. A series of cloned genomic DNA fragments can be screened within 1–2 weeks. The sensitivity of this method is high. Genomic DNA segments of 20 kilobases (kb) or more can be successfully screened in a single transfection by using a set of pooled subclones. This method thus allows the rapid identification of exons in mammalian genomic DNA and should facilitate the isolation of a wide spectrum of genes of significance in physiology and development.

MATERIALS AND METHODS

Cell Culture and Electroporation. COS-7 cells (clonal line A6) were propagated in Dulbecco's modified Eagle's medium supplemented with 10% inactivated fetal calf serum. For transfections, COS-7 cells were grown to 75–85% confluency, trypsinized, collected by centrifugation, and washed in ice-cold phosphate-buffered saline (PBS) in the absence of divalent cations. The washed cells ($\approx 4 \times 10^6$) were then resuspended in cold PBS (0.7 ml) and combined in a pre-cooled electroporation cuvette (0.4-cm chamber; Bio-Rad) with 0.1 ml of PBS containing 1–15 μ g of DNA. After 10 min on ice, the cells were gently resuspended, electroporated [1.2 kV (3 kV/cm); 25 μ F] in a Bio-Rad Gene Pulser, and placed on ice again. After 10 min, the cells were transferred to a tissue culture dish (100 mm) containing 10 ml of prewarmed, pre-equilibrated culture medium.

Vector Construction and Oligonucleotides. pSPL1 was constructed as follows: A 2.7-kilobase-pair (kbp) *Taq* I fragment from pgTat [corresponding to nucleotides 68–2775 of human immunodeficiency virus (HIV) isolate HXB3] (9) was cloned into the *Sal* I site of pBluescript+ (Stratagene). A 2.6-kbp *Bam*HI/*Pst* I fragment was isolated from this construct and used to replace the *Bam*HI/*Eco*RI region of pS β -IVS2 (10), a shuttle vector containing the simian virus 40 (SV40) origin of replication and early region promoter upstream of rabbit β -globin sequences, including β -globin intervening sequence 2 (IVS2). This results in removal of β -globin IVS2 and addition of HIV *tat* intron and flanking exon sequences. The *Eco*RI and *Pst* I sites were removed by blunt-end cloning. The *Bam*HI site in this construct was subsequently removed by *Bam*HI digestion followed by blunt ending with mung bean nuclease. Finally, a *Bam*HI site was inserted into the HIV *tat* intron at the unique *Kpn* I site. Oligonucleotide pairs and the predicted lengths of the PCR products generated by spliced RNA from the vector are as follows: DHAB15, CCAGTGAGGAGAAGTCTGCGG; DHAB14, GTGAGC-CAGGGCATTGGCC (689-bp product); SD2, GTGAACCTG-CACTGTGACAAGC; SA2, ATCTCAGTGGTATTGT-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HIV, human immunodeficiency virus; SV40, simian virus 40.

*To whom reprint requests should be addressed.

GAGC (429-bp product); SD1, CCCGGATCCGCGACGAA-GACCTCTCAAGGC (*Bam*HI cloning site at 5' end); SA1, CCCGTCGACGTCGGGTCCCCTCGGGATTGG (*Sal*I cloning site at 5' end) (102-bp product). The antisense oligonucleotides (DHAB14 and SA2) were used as primers in the first-strand cDNA synthesis reactions.

RNA Isolation, RNA/PCR Amplification, and Cloning. Cytoplasmic RNA was isolated 48–72 hr posttransfection, and first-strand cDNA synthesis was performed as follows: RNA (2.5 or 5 μ g) was added to a reverse transcription solution consisting of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.001% gelatin, 200 μ M dNTPs, and 1 μ M 3' oligonucleotide, and the mixture was heated to 65°C for 5 min. RNasin (3.5 units) (Promega) and Moloney murine leukemia virus reverse transcriptase (200 units) (BRL) were added to the reaction mixture (final vol, 25 μ l), which was then incubated at 42°C for 90–120 min.

The entire reverse transcription reaction was then subjected to PCR amplification in a Thermocycler (Perkin-Elmer/Cetus) using the appropriate oligonucleotide pairs. Thirty-five amplification cycles were routinely used and consisted of 1 min at 94°C, 2 min at 55°C–58°C, and 3 min at 72°C. Products were visualized by staining with ethidium bromide after electrophoresis in 1–1.5% agarose gels.

For cloning, the gel-purified RNA/PCR product was subjected to a second PCR amplification with the internal oligonucleotide pair SD1 and SA1, which flank the vector splice junctions and contain *Bam*HI and *Sal*I cloning sites, respectively. The product from this reaction was gel purified, end-repaired with T4 DNA polymerase (New England Biolabs), digested with *Bam*HI and *Sal*I, and cloned into pBluescript II SK+. Cloned products were sequenced by the dideoxynucleotide chain-termination method (11).

Blot Analysis. Restriction endonuclease-digested genomic DNA clones were electrophoresed through 0.8% or 0.9% agarose gels. RNA/PCR products were electrophoresed through 1–1.5% agarose gels. RNA samples were electrophoresed through a 1% agarose/6% formaldehyde gel and blotted onto a GeneScreenPlus membrane (New England Nuclear). Filters were hybridized by standard procedures (12). DNA probes were radiolabeled to high specific activity by the random primer method (13).

RESULTS

The strategy for exon amplification is outlined in Fig. 1. Vector pSPL1 was designed for insertion, at the *Bam*HI site, of mammalian genomic DNA segments 1–4 kbp long. The insertion site is within an intron from the HIV-1 *tat* gene, whose flanking exons and splice sites were substituted for the second intron of the rabbit β -globin gene. The reporter gene is transcribed by the SV40 early promoter and a polyadenylation signal is derived from SV40. Upon transfection of this plasmid construct into COS-7 cells, RNA transcripts are efficiently generated and the *tat* intron sequences are spliced to produce a polyadenylated cytoplasmic RNA (D.D.C., unpublished data).

When a fragment containing an entire exon with flanking intron sequence in the sense orientation is inserted into the *Bam*HI site of the vector, the exon should be retained in the mature poly(A)⁺ cytoplasmic RNA. As an initial test of the system, fragments of a mouse cosmid clone, MaG#9, known to contain exon sequences of the Na,K-ATPase α_1 -subunit gene (14), were subcloned into pSPL1. A 3.5-kbp *Bgl*II fragment of the cosmid was inserted in sense and antisense orientations into pSPL1, followed by transfection into COS-7 cells. Cytoplasmic RNA preparations derived from the transfectants were analyzed by Northern blotting, using the mouse α_1 -subunit cDNA (15) as a probe. An abundant 2.2-kb RNA species was detected only in cells transfected with the sense

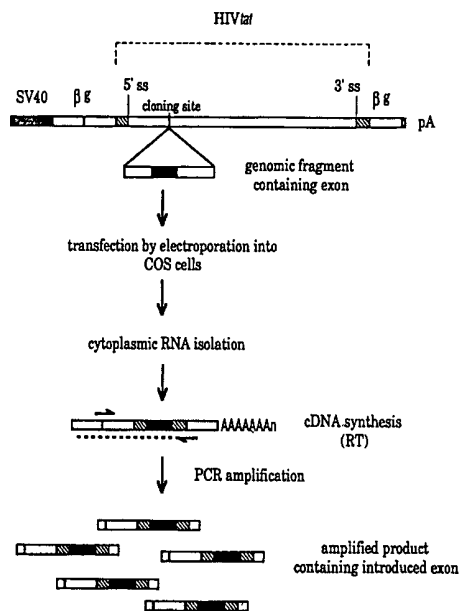


FIG. 1. Structure of pSPL1 and schematic representation of the exon amplification method. A genomic fragment(s) having compatible ends is cloned into the *in vivo* splicing plasmid at a *Bam*HI site created within the HIV *tat* intron and then transiently transfected by electroporation into COS-7 cells. Amplification of the plasmid occurs by virtue of the SV40 origin of replication. High levels of transcription are facilitated by the SV40 promoter, resulting in production of RNA possessing the inserted genomic sequence. If the inserted sequence contains an exon in the proper orientation, processing of the transcript will occur in such a way that the exon is retained in the mature RNA, flanked by HIV *tat* and β -globin exon sequences. Two to three days after transfection, cytoplasmic RNA is isolated from the COS-7 cells and subjected to RNA-based PCR analysis by using oligodeoxynucleotides corresponding to the flanking β -globin sequences. The amplified product contains the introduced exon sequence and can be analyzed either by cloning and sequencing or by direct PCR sequencing. β g, Rabbit β -globin exons; ss, splice site; pA, poly(A) addition recognition sequence; RT, reverse transcription.

construct (Fig. 2A), indicating expression and processing of the transfected sequences.

To isolate spliced exons contained within the vector-derived RNA sequences, we used an RNA-based PCR (RNA/PCR) method, with β -globin-specific oligodeoxynucleotides as primers for the reaction. As expected, oligodeoxynucleotide primers SD2 and SA2 generated an RNA/PCR product of 429 bp from RNA of transfectants with the pSPL1 vector (Fig. 2B, lanes 5 and 8). Analysis of RNA from COS-7 cells transfected with the 3.5-kbp *Bgl*II fragment inserted in the sense orientation into pSPL1 yielded a PCR product of 1.5–1.6 kbp (lane 6). Transfection of a recombinant containing the same fragment in the opposite orientation only yielded the 429-bp PCR product containing vector sequences (lane 7). Hybridization of radiolabeled mouse α_1 -subunit cDNA to blots containing these RNA/PCR products confirmed that sequences derived from the sense construct consist of ATPase exons (Fig. 2C). The length and restriction pattern of the RNA/PCR product derived from sense transfectants are consistent with proper splicing of the six exons of the Na,K-ATPase α_1 -subunit gene contained within this genomic fragment (data not shown).

We performed a more detailed analysis on the RNA/PCR product generated by a 2.8-kbp *Bgl*II fragment from the MaG#9 cosmid. Insertion of this fragment into pSPL1 and transfection yielded a 600-bp RNA/PCR product, which was subsequently cloned and sequenced (data not shown). This product contained

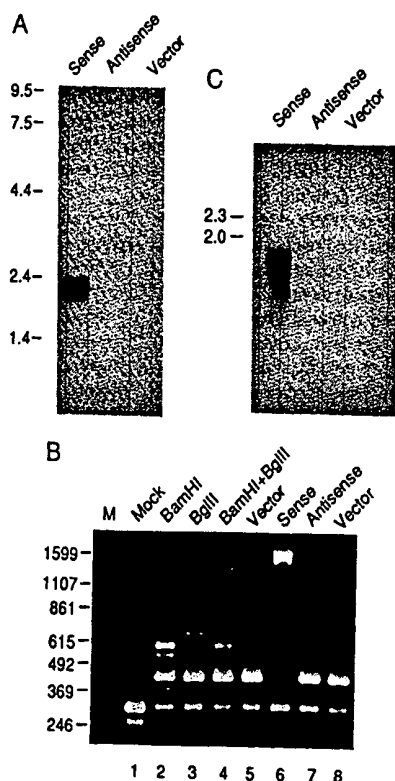


FIG. 2. Exon amplification from a murine Na,K-ATPase α_1 -subunit genomic clone. (A) Northern analysis of RNAs isolated from COS-7 cells transfected with a segment of the Na,K-ATPase α_1 -subunit gene inserted into pSPL1. A 3.5-kbp *Bgl* II fragment from cosmid MaG#9 (12) was subcloned into the *Bam*HI site of pSPL1 in both sense and antisense orientations, and the resulting constructs were transfected into COS-7 cells. RNA preparations (5 μ g) derived from these transfectants were analyzed by Northern blotting with a radiolabeled *Nco* I/*Bam*HI fragment spanning nucleotides 111–705 of the murine Na,K-ATPase α_1 -subunit cDNA used as probe (13). Exposure time for the autoradiograph shown was 4 h. RNA size markers are in kb. (B) The sense and antisense constructs described in A were screened for the presence of exon sequences as described in Fig. 1 and in *Materials and Methods* (lanes 6 and 7). In addition, the entire cosmid MaG#9 was digested with either *Bam*HI or *Bgl* II, or with the combination of these endonucleases followed by shotgun cloning into pSPL1. These constructs were similarly analyzed (lanes 2–5). Oligonucleotides SD2 and SA2 were used as RNA/PCR primers. The resulting RNA/PCR products were visualized by electrophoresis through 1.5% agarose gels and staining with ethidium bromide. The product migrating at 429 bp is derived from splicing occurring between vector 5' and 3' splice sites. This product is absent when a construct containing an exon(s) inserted in the sense orientation is analyzed (lane 6). An \approx 300-bp product is present in all lanes, including mock-transfected (no DNA) cells, indicating that this product is an artifact derived from the COS-7 cell background. A weakly staining \approx 600-bp product is also observed in pSPL1 transfectant products, suggesting that low levels of vector-derived sequences may be amplified. DNA size markers are in bp. (C) Sense and antisense RNA/PCR products from an experiment similar to that described in A were blotted and hybridized to the Na,K-ATPase α_1 -subunit cDNA *Nco* I/*Bam*HI fragment probe. The larger size of the product detected in the sense lane (\approx 1.8 kbp), when compared to the product generated in B, is due to use of the oligonucleotide pair DHAB14 and DHAB15 in the RNA/PCR reaction, which will amplify 689 bp of vector sequence. Exposure time for the autoradiograph shown was 1 h. DNA size markers are in kbp.

exon sequences of the α_1 -subunit cDNA, spanning 171 bp from base pair 125–295. This represents precisely two exons of the

gene, whose sequence and structure have recently been characterized (S.L.G., unpublished data). Thus, accurate processing occurred between *tat* and α_1 -subunit splice recognition sequences, resulting in the removal of the HIV *tat* and ATPase intron sequences and the insertion of ATPase exons in the vector-derived mature RNA.

The above studies demonstrate that, in its simplest form, the *in vivo* splicing selection system can be used to amplify exon sequences from individual segments of genomic DNA. However, in situations in which large regions of a chromosome require analysis in this manner, examination of single fragments would be extremely cumbersome. We therefore tested whether multiple fragments could be analyzed simultaneously. The Na,K-ATPase α_1 -subunit cosmid, MaG#9, was digested separately with *Bam*HI, *Bgl* II, or with the combination of *Bam*HI plus *Bgl* II. Each digest was subsequently "shotgun" cloned into pSPL1. These mixtures of clones were then transfected into COS-7 cells and the resulting RNA was analyzed by RNA/PCR. In this situation, the predominant RNA will contain only sequences from the vector pSPL1, since the majority of genomic fragments contain no exon sequences or are inserted in the antisense orientation. PCR analysis of RNA preparations from cells transfected with shotgun clones of *Bam*HI, *Bgl* II, or *Bam*HI plus *Bgl* II digestions of MaG#9 generated multiple products larger than the 429 bp derived from pSPL1 (Fig. 2B, lanes 2–4). The 600-bp *Bgl* II RNA/PCR product was gel purified, radiolabeled, and directly hybridized to a *Bgl* II restriction digest of MaG#9. Hybridization of this product to the 2.8-kbp *Bgl* II genomic fragment demonstrated that the amplified product was derived from a genomic fragment known to contain an exon (data not shown). These results indicate that in a situation in which the complexity of the genomic DNA is high, exon sequences can still be identified in a single transfection. Interestingly, the 1.6-kbp product detected after transfection with the 3.5-kbp *Bgl* II sense construct was not observed in the *Bgl* II shotgun transfection RNA/PCR product(s). This is most likely due to competition among PCR templates, favoring smaller and more abundant substrates. Also, a weakly staining product migrating at \approx 650 bp was observed in nearly all reaction mixtures containing RNA from plasmid (including pSPL1 alone) transfections and is likely to be artifactual.

To further test the ability of the exon amplification system to screen complex genomic DNA for the presence of exons, genomic clones containing 15–20 kbp of human genomic DNA inserts were analyzed. Each of 12 previously uncharacterized λ phage recombinants containing human genomic DNA, derived from a radiation-reduced human-hamster hybrid cell line containing a segment of human chromosome 19 (J.D.B., unpublished data), was digested with *Bam*HI plus *Bgl* II, shotgun cloned into pSPL1, and transfected into

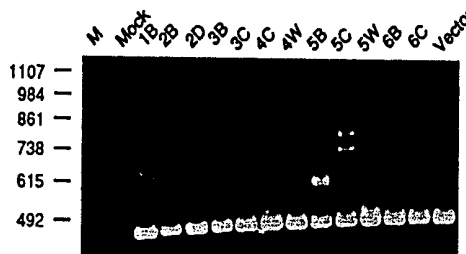


FIG. 3. Exon amplification of anonymous λ genomic clones derived from human chromosome 19. DNA preparations from 12 clones were digested with *Bam*HI and *Bgl* II and analyzed by shotgun cloning for the presence of intact exons, as described in Fig. 2. The previously observed vector-derived \approx 600-bp product is again evident in all pSPL1 transfectant products. DNA size markers are in bp.

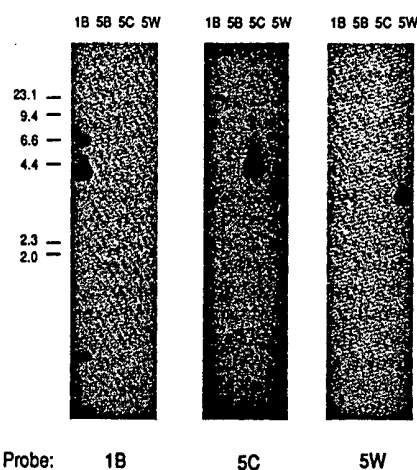


FIG. 4. Hybridization of radiolabeled λ shotgun RNA/PCR products to their corresponding genomic clones. The amplified products from clones 1B (600- to 620-bp doublet), 5C (600-bp product), and 5W (620-bp product) were gel-purified, radiolabeled by the random primer method (11), and hybridized to blots of each λ DNA digested with *Bam*HI plus *Bgl* II. DNA size markers are in kbp.

COS-7 cells. RNA preparations from these transfectants were examined by RNA/PCR (Fig. 3). Six of the 12 amplification reactions (1B, 5B, 5C, 5W, 6B, and 6C) clearly generated products larger than the vector-derived 429-bp product, suggesting that exon sequences are present in each of these clones. The products from 1B (600- to 620-bp doublet), 5C (600-bp product), and 5W (620-bp product) were excised from agarose gels, 32 P-labeled by the random-primer method (13), and hybridized to filters containing blotted DNAs from the original genomic clones. Representative blots are shown in Fig. 4. Each product hybridized only to the genomic DNA segment from which it was derived, indicating that the amplified sequences were not derived from λ phage DNA. The absence of cross-hybridization to other human DNA fragments indicated that the PCR products were essentially free of repetitive sequences. In some cases, two genomic fragments were detected by these probes, suggesting that more than one PCR product was present.

Four of these PCR products were reamplified and cloned by using internal oligonucleotides that correspond to sequences immediately flanking the plasmid splice donor and acceptor sites and that contain cloning sites. Sequence analysis of clones from one of these products, derived from phage

5W, revealed that the RNA/PCR product was derived from an exon of the DNA excision repair gene *ERCC1* (Fig. 5) (16). This gene is located on human chromosome 19 and is known to be present in the human-hamster hybrid cell line from which the genomic clones were derived. A perfect match of the sequence between the HIV *tat* splice junctions and bases 136-247 of the *ERCC1* cDNA sequence (16) indicates that an exon of this gene has been rescued.

We are presently extending the use of exon amplification to uncharacterized regions of the human genome. In preliminary studies, $\approx 70\%$ of cosmid genomic clones (23/33) and 45% of λ phage genomic clones (8/18) have yielded RNA/PCR products containing potential exon sequences. Of these, at least one product appears to contain DNA sequences that are repetitive in nature (a potential false positive), whereas three have demonstrated cross-species sequence conservation. Furthermore, cDNAs corresponding to six other products are currently under characterization (unpublished observations). These results demonstrate the effectiveness of the exon amplification system in the identification of exon sequences in otherwise uncharacterized genomic DNA clones.

DISCUSSION

Exon amplification is a rapid and efficient technique for the identification of expressed DNA sequences in complex mammalian genomes. This method circumvents the laborious characterization of a cloned genomic DNA segment and permits a direct transition to a cDNA. The initial need for appropriate sources of RNA for isolation of cDNA clones is thus also circumvented. The efficacy of exon amplification is clearly demonstrated in this study by the identification and cloning of exons from a cosmid known to contain a portion of the mouse Na,K-ATPase α_1 -subunit gene, as well as exon sequences of the human DNA repair gene, *ERCC1*, from an uncharacterized λ genomic clone. Products of exon amplification could be of particular value in rapidly determining the tissues in which a particular gene is expressed, either by Northern analysis or by *in situ* hybridization. They will also be of use in the isolation of complete cDNAs by library screening procedures or by anchored PCR techniques (17).

Methods related to exon amplification have also been described, including several retroviral systems in which exons can be recovered from genomic DNA inserted into the viral genome (18-25). A recent study by Duyk *et al.* (21) evaluated the use of a retroviral shuttle vector to select for 3' splice sites in random fragments of genomic DNA. The complexity of this system and the length of time required to complete a round of screening are greater than the exon

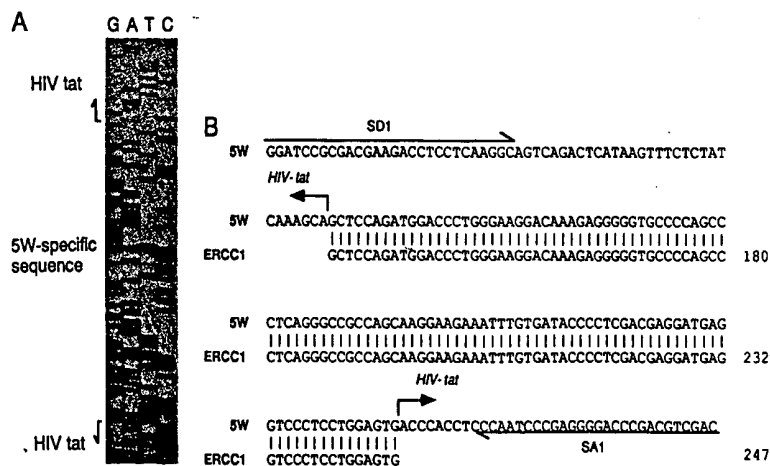


FIG. 5. Sequence analysis of amplified product derived from λ genomic clone 5W. (A) 5W RNA/PCR product was reamplified by using the internal oligonucleotide pair SD1 and SA1 and cloned into pBluescript II SK+. This clone was sequenced by the dideoxynucleotide chain-termination method (11). HIV exon sequences are indicated by arrows. The 5' to 3' sequence is presented top to bottom. (B) Alignment of cloned 5W RNA/PCR product sequence to nucleotides 136-247 of *ERCC1* cDNA (16). Oligonucleotides used for reamplification and HIV *tat* sequences are indicated.

amplification protocol, while the stringency of this retroviral based system is not as great, since the system depends only on the presence of a 3' splice site. Exon amplification will also complement some recently developed methods for isolating transcribed segments of the human genome (24, 25) by permitting removal of intron sequences from cDNAs generated from unprocessed RNA (heterogeneous nuclear RNA) templates. Since these cDNAs represent cloned transcription units, the combination of these approaches should greatly facilitate the cloning of coding sequences.

The nature of the sequence and structure specificity underlying the selection of exons during the splicing of normal nuclear precursor RNAs is not well understood. This specificity is sufficient to screen introns >100,000 nucleotides long in the accurate joining of the flanking exons (26). Experiments suggest that this remarkable specificity is not dictated by the unique nature of the two exons flanking an intron. In fact, all 5' and 3' splice sites are thought to be generically compatible for accurate splicing. This is typified by the accurate splicing of a hybrid intron in which the 5' splice site was derived from a viral exon and the 3' splice site was derived from an exon of the rat preproinsulin gene (18). These results suggest that the exon amplification method should be able to identify most of the exons within a genomic fragment.

There are several potential limitations in the current exon amplification method. First, some types of exons in the screened genomic fragments may not be efficiently spliced into the processed mRNA between the vector-derived exons. This would be the equivalent of exon skipping, which is occasionally observed in the expression of cellular genes (27). The 5' and 3' splice site sequences flanking the pSPL1 vector exons were therefore selected to minimize exon skipping (28). These splice sites are derived from the *rat* exons of HIV-1 and are slowly spliced in both *in vivo* and *in vitro* systems (29). The splice sites of *rat* are compatible for reactions with splice sites from unrelated genes and have been shown to be efficiently spliced to sites flanking the exons of the rat preproinsulin and the rabbit β -globin genes (9, 29). A second potential limitation of the exon amplification process is the selection, in the processed cytoplasmic RNA, of intron sequences from the genomic fragment. These processed intron sequences would probably arise by activation of cryptic splice sites in the inserted sequences. Studies of the splicing of mutant cellular genes suggest that the efficiency of generation of RNA using such cryptic sites would be 1/5th to 1/10th that of normal splice sites (30).

The occurrence of false positives by such mechanisms is an important concern for any selection based on the presence of splice sites. However, there are several indications that they will not constitute a major problem. First, most candidate exon segments selected by the vector do not contain repetitive sequences; thus, random cellular sequences are not appearing frequently in the candidate exon pool. Second, the insertion of genomic fragments, which are thought not to contain exons, have not generated RNA/PCR products. Should false positives occur, it will be possible to distinguish them from true exons by conventional criteria such as cross-species sequence conservation and hybridization discrete mRNA species.

The potential application of exon amplification to large scale screening for transcribed sequences may provide a new approach to genetic mapping. For instance, the construction of transcription maps for large segments of mammalian genomes is technically feasible by this method. Such an approach could provide a powerful adjunct in the fine mapping of the human genome and would enhance the efficiency

with which genes responsible for numerous genetic disorders are identified.

We thank Dr. See-Ying Tam for providing the mouse genomic cosmid clone, MaG#9. This work was supported by Grant HG00299 from the National Institutes of Health and by a grant from the Hereditary Disease Foundation. A.J.B. was supported by Fellowship CA08605 from the National Institutes of Health. J.D.B. was supported by the Muscular Dystrophy Association.

1. Monaco, A. P., Neve, R. L., Colletti-Feener, C., Bertelson, C. J., Kurnit, D. M. & Kunkel, L. M. (1986) *Nature (London)* 323, 646-650.
2. Page, D. C., Mosher, R., Simpson, E. M., Fisher, E. M. C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. & Brown, L. G. (1987) *Cell* 51, 1091-1104.
3. Rommens, J. M., Ianuzzi, M. C., Kerem, B.-S., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, N., Buchwald, M., Riordan, J. R., Tsui, L.-C. & Collins, F. S. (1989) *Science* 245, 1059-1065.
4. Call, K. M., Glaser, T., Ito, C. Y., Buckler, A. J., Pelletier, J., Haber, D. A., Rose, E. A., Kral, A., Yeager, H., Lewis, W. H., Jones, C. & Housman, D. E. (1990) *Cell* 60, 509-520.
5. Fearon, E. R., Cho, K. R., Nigro, J. M., Kern, S. E., Simons, J. W., Ruppert, J. M., Hamilton, S. R., Preisinger, A. C., Thomas, G., Kinzler, K. W. & Vogelstein, B. (1990) *Science* 247, 49-56.
6. Bird, A. (1986) *Nature (London)* 321, 209-213.
7. Chu, G. & Sharp, P. A. (1981) *Nature (London)* 289, 378-382.
8. Horowitz, M., Cepko, C. L. & Sharp, P. A. (1983) *J. Mol. Appl. Genet.* 2, 147-159.
9. Malim, M. M., Hauber, J., Fenrick, R. & Cullen, B. R. (1988) *Nature (London)* 335, 181-183.
10. Buchman, A. & Berg, P. (1988) *Mol. Cell. Biol.* 8, 4395-4405.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
13. Feinberg, A. P. & Vogelstein, B. (1983) *Biochem. Biophys. Res. Commun.* 111, 47-54.
14. Tam, S.-Y., Geissler, E., Graw, S. L. & Housman, D. E. (1990) *Mol. Cell. Biol.* 10, 6619-6623.
15. Kent, R. B., Emanuel, J. R., Ben Neriah, Y., Levenson, R. & Housman, D. E. (1987) *Science* 237, 901-903.
16. van Duin, M., de Wit, J., Odijk, H., Westerveld, A., Yasui, A., Koken, M. H. M., Hoeijmakers, J. H. J. & Bootsma, D. (1986) *Cell* 44, 913-923.
17. Loh, E. Y., Elliott, J. F., Cwirla, S., Lanier, L. L. & Davis, M. M. (1989) *Science* 243, 217-220.
18. Shimotohno, K. & Temin, H. M. (1982) *Nature (London)* 299, 265-268.
19. Sorge, J. & Hughes, S. H. (1982) *J. Mol. Appl. Genet.* 1, 547-559.
20. Cepko, C., Roberts, B. E. & Mulligan, R. C. (1984) *Cell* 37, 1053-1062.
21. Duyk, G. M., Kim, S., Myers, R. M. & Cox, D. R. (1990) *Proc. Natl. Acad. Sci. USA* 87, 8995-8999.
22. Auch, D. & Reth, M. (1990) *Nucleic Acids Res.* 18, 6743-6744.
23. Schwarzbauer, J. E., Mulligan, R. C. & Hynes, R. O. (1987) *Proc. Natl. Acad. Sci. USA* 84, 754-758.
24. Liu, P., Legerski, R. & Siciliano, M. J. (1989) *Science* 246, 813-815.
25. Corbo, L., Maley, J. A., Nelson, D. L. & Caskey, C. T. (1990) *Science* 249, 652-655.
26. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* 55, 1119-1150.
27. Andreadis, A., Gallego, M. E. & Nadal-Ginard, B. (1987) *Annu. Rev. Cell. Biol.* 3, 207-242.
28. Mitchell, P. J., Urlaub, G. & Chasin, L. (1986) *Mol. Cell. Biol.* 6, 1926-1935.
29. Chang, D. D. & Sharp, P. A. (1989) *Cell* 59, 789-795.
30. Wieringa, B., Meyer, F., Reiser, J. & Weissmann, C. (1986) *Nature (London)* 301, 38-43.